Biodiversity Patterns on an Inshore to Offshore Gradient Using Metabarcoding and Barcoding Molecular Tools

Thesis by

Rodrigo Villalobos Vazquez de la Parra

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

King Abdullah University of Science and Technology
Thuwal, Kingdom of Saudi Arabia

© December 2015
Rodrigo Villalobos Vazquez de la Parra
All Rights Reserved
EXAMINATION COMMITTEE APPROVALS FORM

The thesis of Rodrigo Villalobos Vazquez de la Parra is approved by the examination committee.

Committee Chairperson Michael Lee Berumen
Committee Member Carlos Duarte
Committee Member Manuel Aranda
Committee Member Nancy Knowlton
ABSTRACT

Biodiversity Patterns on an Inshore to Offshore Gradient Using Metabarcoding and Barcoding Molecular Tools

*Rodrigo Villalobos Vazquez de la Parra*

It has been estimated that coral reefs shelter 830 000 species. Well-studied biodiversity patterns provide tools for better representation of species in marine protected areas. A cross-shelf gradient in biodiversity exists for fishes, corals, and macroalgae. Here, an inshore to offshore gradient in biodiversity on the Saudi Arabian coast of the Red Sea was sampled using Autonomous Reef Monitoring Structures (ARMS) with barcoding and metabarcoding techniques. It was hypothesized that differences in community structure would be driven by an increase in habitat area. The difference was attributed to the greater accumulation of sediments close to shore that increases the area habitable for sediment dwelling organisms and favors macroalgal cover. Macroalgae are inhabited by a greater number of species than live coral. Only 10% of the sequences of the barcoded fraction and <1% of the metabarcoded fraction had a BLAST hit on the NCBI database with a previously identified species sequence. In addition, the rarefaction curves for all fractions did not plateau. The ARMS community composition changed from inshore to offshore and was significantly correlated with the percentage of algal and bryozoan plate cover. The differences in community composition were related to changes in habitat but not to sediments retrieved from the ARMS.
ACKNOWLEDGEMENTS

I would like to thank Matthieu Leray for his guidance and patience through all the steps of this work, Michael Berumen for his advice and magnificent support through the baseline funding from King Abdullah University of Science and Technology (KAUST), Héloïse Rouzé for her guidance, Nancy Knowlton for allowing me to work in her lab and being part of this thesis committee, Carlos Duarte and Manuel Aranda for being part of this thesis committee, Michael Wooster for his work on the CPCe analysis, The Reef Ecology class of 2013 for deploying the ARMS. Diego Lozano, John Pearman, Francisco Villalobos, Valerio Mazzone, May Roberts, and everyone who helped in the ARMS retrieval. Song He and Tullia Terraneo for their teachings. Maria Peredo for her contribution to the drafting process. Natalia Agudelo for her help with the barcoding protocol. The CMRCL at KAUST for their help during deployment and retrieval of the ARMS. The LAB at the NMNH for allowing the use of their facilities to develop this project. The Reef Ecology Lab at KAUST for their constant support.
TABLE OF CONTENTS

Examination Committee Approval ................................................................. 2
Abstract .................................................................................................................. 3
Acknowledgements .............................................................................................. 4
List of abbreviations............................................................................................ 6
List of illustrations ............................................................................................... 7
List of tables .......................................................................................................... 8
Chapter 1: Introduction ....................................................................................... 9
Chapter 2: Methods ............................................................................................. 14
  2.1 Field sites ..................................................................................................... 14
  2.2 ARMS deployment ....................................................................................... 15
  2.3 ARMS retrieval ........................................................................................... 16
  2.4 Wet laboratory protocol .............................................................................. 18
    2.4.1 ARMS disassembling and sessile fraction processing .......................... 18
    2.4.2 Sieved fractions (106-500 µm; 500 µm-2 mm and >2 mm) ................... 18
  2.4.3 Decantation ............................................................................................. 19
  2.5 Laboratory protocol ..................................................................................... 19
    2.5.1 Barcoding ............................................................................................. 19
    2.5.2 Meta-barcoding ..................................................................................... 21
  2.6 Data analysis ............................................................................................... 22
    2.6.1 OTU assignment ................................................................................... 22
    2.6.2 Alpha and betadiversity-barcoding ....................................................... 23
    2.6.3 Alpha and beta diversity-metabarcoding ............................................. 24
    2.6.4 Plate coverage analysis ......................................................................... 24
    2.6.5 OTUs distribution analysis .................................................................... 25
Chapter 3: Results ............................................................................................... 26
  3.1 OTU clustering ............................................................................................. 26
  3.2 Alpha Diversity ............................................................................................ 27
  3.3 Beta Diversity .............................................................................................. 31
Chapter 4: Discussion ......................................................................................... 36
Chapter 5: Conclusions ....................................................................................... 40
References ............................................................................................................ 41
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMS</td>
<td>Autonomous Reef Monitoring Structures</td>
</tr>
<tr>
<td>CPCe</td>
<td>Coral Point Count with Excel extensions</td>
</tr>
<tr>
<td>KAUST</td>
<td>King Abdullah University of Science and Technology</td>
</tr>
<tr>
<td>OTU</td>
<td>operational taxonomical unit</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>PCoA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>UPGMA</td>
<td>Unweighted pair group method with arithmetic mean</td>
</tr>
</tbody>
</table>
LIST OF ILLUSTRATIONS

- Figure 1. Location of the ARMS deployment by replicates of three 16
- Figure 2. Pictures of the locations of the 12 replicate ARMS in the central Saudi Arabian Red Sea 17
- Figure 3. Step function analysis measuring the different number of OTUs retrieved with each dissimilarity threshold used on MOTHUR. 27
- Figure 4. Rank abundance plot 28
- Figure 5. Rarefaction curves for barcoded and metabarcoded fractions 30
- Figure 6. UPGMA tree showing beta diversity among samples of the >2mm fraction. 34
- Figure 7. 3D Principal component analysis of species composition and its abundance of each sample 34
- Figure 8. 2D Principal component analysis of species composition and its abundance of each sample with confidence ellipsoids using jackknife 35
LIST OF TABLES

Table 1. Three most common families of corals on Fsar, Al Fahal South, Shib Nazar

Table 2. Number of organisms collected per site of the three most common phyla of the >2mm fraction

Table 3. Percentage cover of most abundant groups on each sample
Chapter 1: Introduction

Biodiversity patterns have been intriguing to scientists since the 1950’s (Hutchinson 1959). Ecosystems with high biodiversity are more likely to maintain their ecosystem services compared to low biodiversity ecosystems, unless complete functional groups are removed from the system (Bellwood et al. 2003; Hughes et al. 2005). Understanding biodiversity patterns helps predict the response of species diversity and community composition to environmental change (Gaston 2000).

The estimated total number of species in the ocean is 300,000 – 2,200,000 species with 220,000 of them already described (Bouchet 2006; Costello & Wilson 2011; Mora et al. 2011; Appeltans et al. 2012). Coral reefs shelter more than one third of the ocean’s biodiversity, with an estimate of 830,000 species. Out of these 830,000 only 70,000 are named species (Fisher et al. 2015). Coral reefs occupy less than 0.1% of the ocean floor (Spalding & Grenfell 1997).

Shallow water habitats have been more studied than deeper and pelagic ecosystems. However in some areas, like the Red Sea, the study of coral reefs has been minimal compared to systems such as Australia’s Great Barrier Reef (Berumen et al. 2013).

The cost and time of the use of traditional approaches to be able to describe all species has been estimated to take centuries and 270 billion dollars (Carbayo & Marques 2011; Mora et al. 2011). However, with the use of operational
taxonomic units (OTU) we can reveal biodiversity patterns without the need of identifying or describing species. An OTU is a group of organisms with morphological or genetic similarities considered to be within the boundaries of a single species. The use of DNA sequence similarities, in contrast to morphological based OTUs, facilitates the comparison of studies (Plaisance et al. 2011b; Ratnasingham & Hebert 2013). Such genetic tools should not be seen as a replacement for traditional taxonomy, but rather as a way to measure biodiversity while traditional taxonomy catches up with the overwhelming task of describing all marine species.

The use of Autonomous Reef Monitoring Structures (ARMS) to measure biodiversity patterns and document species richness of cryptic communities has been used more and more over the last decade (Knowlton et al. 2010; Plaisance et al. 2011a; Plaisance et al. 2011b; Leray & Knowlton 2015). ARMS are deployed on the seafloor for 6 months to 3 years. During that period new recruits colonize the ARMS. The percentage cover of the plates of the ARMS will vary with time and location. Plate cover percentage could be used to compare sampled locations. The ARMS are not able to collect all biodiversity on coral reefs, but are used as a standardized and cost-effective method to record trends in cryptic biodiversity (Knowlton et al. 2010). Alongside the ARMS, small dead colonies of branching corals, usually of the genus *Pocillopora*, are used to study cryptic biodiversity. However, the complexity of corals presents a major challenge for standardization (Plaisance et al. 2009; Knowlton et al. 2010). Compared to the *Pocillopora* sampling method, the ARMS can be deployed on any kind of
benthic ecosystem and does not destroy the natural habitat of the cryptic community (Knowlton et al. 2010).

One of the first accepted biodiversity patterns in ecology is that species diversity will increase with an increase in area sampled (Williams 1964). This pattern occurs because an increase in area increases the opportunity to sample species with low populations as well as habitats hosting un-sampled species (Rosenzweig 1995; Allouche et al. 2012). A second pattern of biodiversity states that a larger variety of habitats will lead to an increase in species diversity (Macarthur 1958). Within a taxon, each individual species will prefer a specific habitat, or niche, within a sampling area (Rosenzweig 1995; Allouche et al. 2012). For example, if only the exposed area of a rocky shore is sampled, it is possible that no crabs will be collected, and the sample will be full with barnacles. Even though crabs are present, the sample area might not be large enough to gather an accurate representation of the biodiversity. Area sampled and habitat heterogeneity have an unimodal relationship with biodiversity when considered together (Allouche et al. 2012). The area of each type of habitat will decrease with an increase in the number of habitats within the area sampled. Latitudinal and longitudinal changes throughout an ecosystem will lead to differences in species diversity as well. An increase with latitude will have a decrease in species diversity for most taxa, although some taxa will have their peak of diversity at mid latitudes and sometimes even at higher latitudes rather than the tropics (Rosenzweig 1995; Gray 1997; Lomolino et al. 2010; Leray & Knowlton 2015). On coral reefs, species diversity decreases with longitudinal distance from
the Coral Triangle (Veron 1995; Gray 1997; Veron et al. 2011). Disturbances in an ecosystem regulate diversity, with maximum diversity predicted to occur at intermediate levels of disturbance (Connell 1978; Huston 1985; Rosenzweig 1995). Consumers (or predators) and productivity also have a unimodal relationship with species diversity, because consumers act in a similar way to disturbances, controlling the population of the dominant species and preventing competitive exclusion (Huston 1985; Cornell & Karlson 2000; Worm et al. 2002). Productivity can have also a positive relationship with species richness (Cornell & Karlson 2000) but with excessive nutrients causing lower diversity (Irigoien et al. 2004).

In the past, coral reef biodiversity patterns in the Red Sea have been compared to localities in the Indian Ocean and Arabic Sea (Sheppard 1998). Cross-shelf gradients of biodiversity have been studied in taxa such as corals, fish, and macroalgae (De’Ath & Fabricus, 2010; Malcolm et al. 2010). De’ath and Fabricius (2010) found a decrease in species richness of heterotrophic corals from inshore to offshore reefs. Macroalgae also decreases in abundance from inshore to offshore reefs. The macroalgae is benefitted by the decrease in water quality, while the heterotrophic coral diversity has the opposite relationship (De’ath & Fabricius 2010). Malcolm et al. (2010), aiming to provide information for future conservation planning, found an increase in species diversity in fishes from inshore to offshore reefs. Well-known biodiversity patterns and species distributions will provide tools for broader representation of biodiversity in marine protected areas (Ward et al. 1999; Cowling et al. 2003).
In the present study, we assessed the biodiversity of organisms that colonized the ARMS on inshore, midshelf, and offshore reefs using molecular tools. Sedimentation rate has a negative relationship with coral species richness, but a positive relationship with macroalgae abundance (Rogers 1990; De'ath & Fabricius 2010). Algal turfs shelter a high number of the mobile invertebrates (Milne & Griffiths 2014). The sediment deposition provides an increase in sediment dwellers habitat area, and a change in benthic cover distribution, which may benefit the diversity of the ARMS community (Rogers 1990; De'ath & Fabricius 2010). Therefore it is expected that the species diversity will decrease from inshore to offshore. If an increase of species diversity from inshore to offshore is observed, it would indicate that other factors are changing the community structure on an inshore to offshore gradient. It is expected to find a gradual change in biodiversity and community structure of the crypto-fauna from inshore to offshore reefs on the Saudi Arabian coast of the Red Sea. Results can contribute to understanding the undescribed proportion of species the reefs of the coast of Saudi Arabia on the Red Sea compared to other locations.
Chapter 2: Methods

2.1 Field sites

To test for a biodiversity gradient from inshore to offshore reefs, the locations were chosen to include one inshore reef, two locations on a midshelf reef, and one offshore reef.

The inshore reef, Fsar (22° 13.857’ N, 39° 1.7436’ E), is sheltered from the waves and has a shallow bottom. The midshelf locations were chosen on the reef Al Fahal that presents different characteristics on the north side (22° 18.089’ N, 38° 57.603’ E) versus the south side (22° 13.857’ N, 39° 1.7436’ E). The south side is more sheltered and is farther from deep water than the north side. The offshore reef Shib Nazar has a steep slope that rapidly reaches deep water. It is also more exposed to wave action than the inshore and midshelf locations (Fig. 1).

The most abundant coral family in 2011 on Fsar was Poritidae, while on Al Fahal South and Shib Nazar the dominant coral family was Pocilliporidae (Furby et al. 2013). There is no information available for Al Fahal N. Regionally, the fish biomass decreases with the proximity to a port, and on the KAUST reefs the distance from port increases with distance from shore (Kattan 2014).
Table 1. Three most common families of corals on Fsar, Al Fahal South, Shib Nazar surveyed with a line-intercept method (Furby et al. 2013).

<table>
<thead>
<tr>
<th>Reef</th>
<th>Most abundant families of corals from left to right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fsar</td>
<td>Poritidae, Faviidae, Fungiidae</td>
</tr>
<tr>
<td>Al Fahal S</td>
<td>Pocilliporidae, Faviidae, Acroporidae</td>
</tr>
<tr>
<td>Shib Nazar</td>
<td>Pocilliporidae, Faviidae, Acroporidae</td>
</tr>
</tbody>
</table>

2.2 ARMS deployment

Three ARMS were deployed in 10 m depth at each of four locations on the exposed side of the reefs outside King Abdullah University of Science and Technology (KAUST; Fig. 2). Each ARMS consists of nine 22.5 x 22.5 cm plates with a space of 1.27 cm between them. The flow of water inside the ARMS is obstructed placing rectangular bars placed in an x shape in alternating spaces.
Figure 1. Location of the ARMS deployment by replicates of three; one inshore reef (Fsar), two locations on a midshelf reef (Al Fahal South and Al Fahal North), and one offshore reef (Shib Nazar)

2.3 ARMS retrieval

The ARMS were retrieved after approximately two years underwater, being deployed in February 2013 and recovered in May 2015.

All the equipment for ARMS retrieval and processing was soaked 10 minutes in 10% Sodium hypochlorite and rinsed with tap water.

The ARMS were retrieved with a prefabricated plastic box covered with a 100 µm mesh to avoid the loss of specimens. The plastic box was positioned on top of the ARMS with bungee cords creating pressure to seal the plastic box with the bottom of the ARMS. After carefully being brought to the surface, the ARMS were placed in a large plastic heavy-duty box (102 L capacity) filled with in situ filtered seawater to 45 µm. Battery-powered air pumps were used to aerate the water inside the boxes during transit back to the laboratory.
Figure 2. Pictures of the locations of the 12 replicate ARMS in the central Saudi Arabian Red Sea. Photos were taken prior to their recovery; A-C) the inshore site, Fsar; D-E) a midshelf site, Al-Fahal South; F-H) another midshelf site, Al-Fahal North; I-L) an offshore site, Shib Nazar. Photo credits: Matthieu Leray
2.4 Wet laboratory protocol

2.4.1 ARMS disassembly and sessile fraction processing

The ARMS were disassembled within the heavy-duty box with battery-powered air pumps inside. Each plate of the ARMS was gently brushed with a small brush inside the 102 L container to remove all mobile animals. Each plate was put on a labeled tray containing 1 µm-filtered seawater and a battery-powered air pump, and was photographed for a percentage cover analysis. Each plate was scraped and immediately filtered on a 45 µm mesh to remove the excess of water. The sessile fraction was then homogenized using a blender to create an initial sample. The product of the blending was filtered again to remove the excess of water and mucus. Three 50 mL tubes were filled with 25 mL of the sessile fraction. Two of them were preserved using DMSO and the third using 80% ethanol. The rest was kept in a resalable plastic bag and frozen at -40°C.

2.4.2 Sieved fractions (106-500 µm; 500 µm-2 mm, and >2 mm)

The water from the 102 L container was poured through a 2 mm sieve and collected in another previously bleached 102 L container. The animals collected from this fraction were placed on a labeled tray with 1 µm-filtered seawater and a battery-powered air pump. The water was then poured again through a 500 µm sieve and consecutively through a 106 µm sieve. These fractions were also each placed in a tray with 1 µm-filtered seawater and a battery-powered air pump. The >2 mm animals were sorted out manually and placed into individual flasks with a battery-powered air pump, sorted by broad taxonomic groups. Each organism
was photographed for future identification and preserved in 80% ethanol. The sediments and organisms collected with the 106 µm and the 500 µm sieves were filtered through a 45 µm mesh and then preserved in 96% ethanol.

2.4.3 Decantation

The 106-500 µm and the 500 µm-2 mm fractions were separated from the sediments through the same procedure. Each sample was moved to a bigger container, either a 1 L or 2 L. The container was filled three quarters with reverse osmosis water and moved to separate the animals from the sediment. Then the water was poured through a 45 µm sieve. The organisms were collected from the sieve with a sterile disposable spatula and placed in a 50 mL falcon tube. The product was weighed, evenly mixed, and divided in two. One half was crushed and homogenized using a mortar and pestle for 2 min and preserved in 96% ethanol. The other half was kept as an archive in 80% ethanol. The sediments remaining on the bottom of the bottle were weighed and preserved in 96% ethanol.

2.5 Laboratory protocol

2.5.1 Barcoding

Each group of organisms was tissue sampled aiming to get a sufficient amount of muscle and preserve the organism in the best condition possible.

Turbellarians, sipunculids, polychaetes, and holothurians. One small part from the middle of the organism was taken, avoiding the anterior and posterior end.
Mollusks. Muscle from the foot was taken. For organisms inside the shell, the shell was broken gently and the pieces recovered and stored in 80% alcohol.

Crabs. A leg of each crab was removed opening a space where then muscle was extracted from the thorax. For small crabs all the legs from one side were removed, obtaining tissue from the thorax as well.

Shrimps. The muscle in the middle part of the abdomen of each shrimp was taken, avoiding the exoskeleton.

Other crustaceans. A sample of muscle was extracted from the abdomen avoiding taking any leg and exoskeleton.

Ophiuroids and sea urchins. Tissue from around the mouth was gently extracted. In case the ophiuroid was too small also the distal segments of one arm were also smashed and added to the extraction.

Tissue sampled from each organism was placed in a 96-well Costar plate, followed by phenol DNA extraction on the AutoGeneprep 965 from AutoGen. The DNA extract was eluted and a 658bp region of the mitochondrial COI gene was amplified through PCR with the barcode primers jgLCO and jgHCO (Geller et al. 2013). The PCR mix was as follows: 6.6 µL nuclease free water, 10 µL Promega GoTaq G2 Hot Start Master Mix, 0.2 BSAµL (20mg/mL), 0.6 µL jgLCO (10µM), 0.6 µL jgHCO (10µM), and 1 µL eluted DNA. For the PCR thermal cycler the temperatures and times were 95°C for 5 min; four cycles at 94°C for 30 s, 50°C for 45 s, and 72°C for 1 min; thirty-four cycles at 94°C for 30 s, 45°C for 45 s, and...
72°C for 1 min; 72°C for 8 min. An electrophoresis gel was done to test the PCR product. The components of the gel were as follows: 1.5% Agarose LE in sodium hydroxide buffer with 2 µL GelRed from Biotium. On each well, 1 µL loading dye and 1 µL PCR product was placed. The gel was placed at 80 V, 120 mA for 14 min. The unsuccessful PCR products were discarded and the PCR was repeated with the dgLCO/dgHCO primer combination (Meyer 2003). The PCR product was then purified using ExoSAP-IT from Affymetrix and sequenced in both directions using Sanger sequencing on the Sanger ABI 3730 capillary platform.

2.5.2 Meta-barcoding

The DNA from the 10 g of the homogenized sessile fraction and the homogenized 106-500 µm and 500 µm-2 mm fractions were extracted using the MO-BIO Powermax Soil DNA Isolation Kit following the manufacturer’s protocol in all but the first bead-beating step, which was changed by adding proteinase K to the power bead solution and shaking at 56°C overnight as suggested by Leray and Knowlton (2015). Following a DNA purification step using the Mo-BIO Powerclean DNA Clean-Up Kit, a hierarchical tagging approach was used for PCR amplification, using seven tailed PCR primers. The samples were sequenced with high throughput sequencing on the Miseq Illumina platform. Three PCRs using the mlCOIintF and tailed-jgHCO primer combination with the same conditions were made on each sample of the scraped, 106-500 µm, and 500 µm-2 mm fractions to amplify 313 bp of the COI mitochondrial gene. The mixture used in the PCRs was as follows: 1.4 µL of dNTP (10mM), 0.4 µL of
Clontech Advantage 2 Polymerase Mix, 10 µL reaction with 1 µL of tailed-mlLCO (10 µM), 10 µL reaction with 1 µL of tailed-jgHCO (10 µM), 2 µL of the PCR buffer Clontech Advantage 2, and 1 µL of the purified DNA (10 ng). The use of the primer set mlCOIintF and tailed-jgHCO amplify 91% of metazoan diversity (Leray et al. 2013).

2.6 Data analysis

2.6.1 Operational taxonomic unit (OTU) assignment

Sequences were grouped together in OTUs, which we assumed to be a distinct evolutionary lineage. The forward and reverse sequences from the barcoded fraction were assembled and edited using Genious (Biomatters). The 5’ and 3’ ends of each sequence with more than a 5% chance of an error per base were trimmed. The sequences were considered not trustable and discarded if they contained one or more stop codons or more than three mistakes leading to different amino acid translation in the sequencing. The sequences were aligned using a ClustalW alignment with a high open gap and gap extent cost using Genious (Biomatters). Sequences from the barcoded fraction were clustered into OTUs using MOTHUR (Schloss et al. 2009) and sequences from all fractions were clustered into OTUs using an unsupervised Bayesian clustering method CROP (Hao et al. 2011) as a comparison method. Lower bound variance was set to 3 and upper bound variance to 4 on CROP as in Leray et al. (2013). The representative sequences output by CROP were BLAST in ViroBLAST (Deng et al. 2007) using the NCBI nucleotide database to obtain a taxonomic match. A
threshold of 97% assignment was used following Leray and Knowlton (2015). In addition, the photographs of each OTU were identified to the lowest taxonomic level possible using pictures from organisms previously collected in the Red Sea and identified by experts. The clusters of OTUs obtained from CROP for all the fractions were transformed into an OTU table to do the downstream alpha and beta analysis in QIIME (Caporaso et al. 2010).

2.6.2 Alpha and beta diversity-barcoding

Rank abundance curves of number of individuals per OTU for each sample were created to observe the evenness of each sample. Heatmaps with log transformed numbers of COI amplicons within each sample per taxonomic group were plotted to observe the differential taxa distribution among samples. Alpha diversity was determined using the Chao1 parameter and number of species per sample, considering each OTU as a different species. Rarefaction curves were plotted to observe if the sample effort was enough to sample most of the cryptic species. A qualitative, binary Jaccard (Jaccard 1901), and a quantitative, Bray Curtis (Bray & Curtis 1957), method were used to measure beta-diversity. The Jaccard method was chosen because it is widely used, allowing comparison between studies; however it does not consider the evenness of the community (Sheppard 1998; Marcia Barbosa 2015). The Bray-Curtis was chosen because it provides a comparative measure between communities considering both abundance and species diversity, giving a reliable measure of comparison between communities (Clarke 1993).
2.6.3 Alpha and beta diversity - metabarcoding

OTUs with only one sequence were removed from the analysis, as they are likely to be sequencing errors. To avoid errors from the alpha and beta diversity measures being calculated from unequal numbers of reads per sample, the larger samples were scaled to the number of sequences of the smallest sample in the dataset using the single_rarefaction.py. 2D and 3D principal component analysis (PCoA) as well as an unweighted pair group method with arithmetic mean (UPGMA) trees were built to observe differences between reefs, samples, and fractions. The branches of the UPGMA tree were weighted using jackknife. The differences between groups of samples were tested with a PERMANOVA. Alpha and beta diversity analyses were performed using QIIME.

2.6.4 Plate coverage analysis

The percentage of plate cover was measured with a point count approach using Coral Point Count with Excel extensions (CPCe). Each plate was analyzed on the bottom and top (upper and lower surfaces) side at the phylum level. A grid of 15x15 points was placed over each picture of the plate for all the nine plates of each ARMS.
2.6.5 OTUs distribution analysis

To test for changes in the community structure of the >2mm fraction related to habitat type, a nonparametric statistical method, adonis, was implemented using QIIME. Habitat type was measured as the percentage of plate covered by each phylum and in total, from the results obtained on the CPCe analysis and as sediment content per sample for the 106-500 µm and 500 µm-2 mm fractions.
Chapter 3: Results

3.1 OTU clustering

For the whole set of the >2 mm fractions the program MOTHUR with a 0.05 dissimilarity threshold retrieved 174 OTUs while CROP retrieved 167 OTUs. Using MOTHUR, between 0.05 and 0.11 dissimilarity thresholds, a plateau was observed for this dataset (Fig 3). Below 0.04 the slope is notably steeper, meaning that there is a considerable difference in OTU numbers from using close dissimilarity thresholds.

For the metabarcoded fraction, CROP retrieved 6029 OTUs, and after filtering out the OTUs with only one sequence on QIIME, 4463 OTUs were left. The reads per sample were further diminished to match each sample with the sample that had the minimum number, having 4253 OTUs.
3.2 Alpha Diversity

In the barcoded and metabarcoded fractions 70% and 45% of the OTUs respectively, were represented by less than 3 sequences in each location. The samples were dominated by a few species (Fig. 4).
Figure 4. Rank abundance plot; a) Absolute abundance against species rank of the barcoded fraction and b) metabarcoded fractions; the evenness of each sample is represented by the slope of the curve.

For the >2 mm fraction, Fsar and Al Fahal N reefs were the sites with the fewest OTUs (54 and 44 OTUs respectively). Al Fahal S and Shib Nazar had 76 and 68 OTUs respectively. The Chao1 index showed Fsar with the lowest alpha diversity (85.9 Chao1). Al Fahal S, Al Fahal N, and Shib Nazar had 117.35, 98.2, and 154.1 Chao1 estimates respectively. The most abundant and diverse group in all locations was Arthropoda (114 OTUs and 513 sequences). Arthropoda, Mollusca, and Annelida were the groups with more OTUs at all sites (Table 2). Platyhelminthes was found only in Al Fahal N and Sipuncula only in Al Fahal S. Phyla were not determined for the metabarcoded fractions.
Table 2. Number of organisms and species collected per site for the three most common phyla of the >2mm fraction

<table>
<thead>
<tr>
<th></th>
<th>Arthropoda #Sequences</th>
<th>OTUs</th>
<th>Mollusca #Sequences</th>
<th>OTUs</th>
<th>Annelida #Sequences</th>
<th>OTUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fsar</td>
<td>105</td>
<td>21</td>
<td>32</td>
<td>12</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Al Fahal S</td>
<td>174</td>
<td>34</td>
<td>42</td>
<td>20</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Al Fahal N</td>
<td>89</td>
<td>19</td>
<td>18</td>
<td>11</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Shib Nazar</td>
<td>145</td>
<td>40</td>
<td>24</td>
<td>16</td>
<td>12</td>
<td>6</td>
</tr>
</tbody>
</table>

None of the rarefaction curves of the barcoded and metabarcoded fractions reached a plateau (Fig. 5). The rarefaction curves also did not plateau when all the metabarcoded fractions were considered together. The fraction with the most OTUs was the 106-500 µm fraction. Among the 167 OTUs on the >2 mm fraction, only 26 OTUs (15.6%) had a match in sequences from the National Center for Biotechnology Information (NCBI) nucleotide database with 97% similarity or more. Of these, 17 OTUs (10.2%) had a match with a previously identified species on the NCBI nucleotide database. In the metabarcoded fraction, 57 OTUs out of 4463 (1.3%) had a match with a sequence on the NCBI database and 40 OTUs (0.9%) of these were previously identified to species level.
Figure 5. Rarefaction curves for barcoded and metabarcoded fractions; A) >2 mm fraction with observed species; B) >2 mm fraction with Chao1 parameter; C)
all metabarcoded fractions with observed species; D) all metabarcoded fractions with Chao1; From A-D: In red Fsar; in orange Al Fahal South; in blue Al Fahal North; in green Shib Nazar; The circle encloses the 106-500 µm fraction.

3.3 Beta diversity

Fsar samples were neatly grouped together in the barcoded fraction through an UPGMA and supported by jackknife subsampling on both Jaccard and Bray Curtis. The midshelf reef Al Fahal North clustered together on the Bray Curtis UPGMA tree with a poor support on the jackknife subsampling. On the Jaccard UPGMA tree Shib Nazar was clustered together as a unique group supported with more than 75% on the jackknifing subsampling (Fig. 6A). The distinctiveness of the inshore reef Fsar is also shown in the 3D PCoA (Figs. 7A, 7B) and in 2D PCoA (Figs. 8A, 8B), where for the barcoded fraction the axes explained 19.73% and 29.72% using Jaccard and Bray Curtis respectively (Fig. 8). Al Fahal North and South occupied partially overlapping space on these axes, indicating similar community composition. The PERMANOVA test showed a significant difference between barcoded samples (Jaccard p-value 0.001; Bray Curtis p-value 0.002).

The combined metabarcoded fractions did not cluster together by location on the UPGMA trees (Figs. 7C, 7D). The PCoA in a 3D space showed a slight change in community from the reefs offshore towards those inshore and protected (Figs. 7C, 7D) with some overlapping, using both Jaccard and Bray Curtis. The Jaccard PCoA in a 2D space for the metabarcoded fraction on an axis explaining 14.65% of the variation also showed a slight change in community composition from
inshore to offshore (Fig. 8C). The Bray Curtis PCoA on a 2D space (Fig 8D), on an axis explaining 40.45% of the variation, indicates similarities between the North and South locations on the midshelf reef and the inshore reef communities with the offshore reef community. However, considering all axes, the inshore reef and the North and South locations on the midshelf reef occupy a different space on the Bray Curtis 2D PCoA. The PERMANOVA test showed a significant difference between samples on the metabarcoded fractions (Jaccard p-value 0.002; Bray Curtis p-value 0.004).

The nonparametric analysis tested on the >2mm fraction indicated a significant relationship between the percentage covered on each plate by algae (p-value 0.005) and bryozoans (p-value 0.028) and the changes in the community of the >2mm fraction. The coral cover of plates was close to the significance threshold for this comparison (p-value 0.059) (Table 3). The sediment content did not show a significance relationship with changes in community composition of either fraction.

Table 3. Percentage cover of most abundant groups on each sample; FR1, FR2 and FR3 samples belong to Fsar reef; FS1, FS2 and FS3 to Al Fahal S; FN1, FN2 and FN3 to Al Fahal N; SN1, SN2 and SN3 to Shib Nazar

<table>
<thead>
<tr>
<th></th>
<th>FR1</th>
<th>FR2</th>
<th>FR3</th>
<th>FS1</th>
<th>FS2</th>
<th>FS3</th>
<th>FN1</th>
<th>FN2</th>
<th>FN3</th>
<th>SN1</th>
<th>SN2</th>
<th>SN3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algae</td>
<td>32</td>
<td>28</td>
<td>31</td>
<td>26</td>
<td>27</td>
<td>29</td>
<td>27</td>
<td>26</td>
<td>26</td>
<td>23</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>Bryozoan</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>19</td>
<td>21</td>
<td>19</td>
<td>21</td>
<td>20</td>
<td>20</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>Porifera</td>
<td>17</td>
<td>17</td>
<td>22</td>
<td>24</td>
<td>31</td>
<td>18</td>
<td>20</td>
<td>18</td>
<td>21</td>
<td>18</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>Bare</td>
<td>11</td>
<td>13</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>15</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Coral</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Plate Cover</td>
<td>89</td>
<td>87</td>
<td>90</td>
<td>93</td>
<td>94</td>
<td>93</td>
<td>90</td>
<td>93</td>
<td>91</td>
<td>91</td>
<td>85</td>
<td>89</td>
</tr>
</tbody>
</table>
Figure 6. UPGMA tree showing beta diversity among samples of the >2mm fraction with branch support calculated using jackknifing; red branches 70-100% support, yellow branches 50-75% support; green branches 25-50% support and blue branches <25% support A) Jaccard UPGMA hierarchical clustering tree; B) Bray Curtis UPGMA hierarchical clustering tree.

Figure 7. 3D Principal component analysis of species composition and its abundance of each sample; the ellipsoids around the dots represent the confidence using jackknife; A) Barcoded fraction Jaccard; B) barcoded fraction
Bray Curtis; C) Metabarcoded fraction Jaccard; D) metabarcoded fraction Bray Curtis. In red Fsar; in orange Al Fahal S; in blue Al Fahal N; in green Shib Nazar

Figure 8. 2D Principal component analysis of species composition and its abundance of each sample with confidence ellipsoids using jackknife; A) Barcoded fraction Jaccard; B) Barcoded fraction Bray Curtis; C) Metabarcoded fraction Jaccard; D) Metabarcoded fraction Bray Curtis
Chapter 4: Discussion

The low hits on blast percentage (15% on the barcode fraction and 1% on the metabarcoded fractions) with NCBI database indicate the low level of knowledge of the Red Sea. In addition the sampling was not exhaustive enough to have a representation of all the cryptic species on coral reefs of the Saudi Arabian coast (Berumen et al. 2013). The results from the 3DPCoA and 2DPCoA (Figs. 8 and 9 respectively) show a change in community structure from an inshore to offshore gradient, which might be due to the change in type of habitat, if we consider the type of plate cover as the habitat for the crypto fauna (Rosenzweig 1995). The cryptofauna community of reefs seems to have a low evenness in several locations worldwide (Plaisance et al. 2009; Plaisance et al. 2011b; Leray & Knowlton 2015).

The Red Sea shelters a high diversity and most of it is unknown (Veron 1995; Berumen et al. 2013). Here only 17 OTUs out of the 167 of the barcoded fraction and 40 OTUs out of the 4463 on the metabarcoded fractions got a hit with an identified species when blasting them on the NCBI database. Among the metabarcoded fractions, the 106-500 µm fraction was the most diverse fraction in this study. Also rarefaction curves for all the fractions did not plateau. Leray and Knowlton (2015) used the same protocol as in this study and found that on oyster reefs of Florida and Virginia over half of the species from the >2mm fraction had a match with a sequence on the NCBI database. The reef cryptofauna of other
locations also showed rarefaction curves that did not plateau (Plaisance et al. 2011a; Plaisance et al. 2011b; Leray & Knowlton 2015). More research is needed to describe the full extend of cryptofauna species on the Saudi Arabian Red Sea coral reefs.

There is a change in community composition from inshore to offshore reefs due a change in abundance on the most common groups. The abundance of the most common group of the barcoded fraction, the Galatheidae, increased from inshore to offshore. The second most abundant group from the barcoded fraction, the Palaemonidae, had the opposite change in abundance. A difference in habitat heterogeneity and habitat area might be affecting these groups and therefore the whole community structure (Macarthur 1958; Rosenzweig 1995; Bellwood & Hughes 2001). Our results showed that algae coverage and bryozoan coverage had a significant influence on the ARMS community composition. They might be acting as a habitat for some of the species collected on the ARMS. The inshore to offshore gradient on species distribution had been observed also on corals, macroalgae, and fish (De’ath & Fabricius 2010; Malcolm et al. 2010). (Rogers 1990; De’ath & Fabricius 2010) described a change in community structure with different sedimentation rates. Here the sediment concentration did not show any relationship with the changes in community structure. The PCoA of the metabarcoded fractions also showed a change from inshore to offshore in the ARMS community structure.
All the reefs near KAUST might be connected through larval dispersal for the ARMS community, which in addition to successful settlement of the larvae could explain the proximity of the samples from different reefs on the PCoA, on the gradient inshore to offshore. The KAUST reefs have almost no self-recruitment in clownfish, although the reefs could be physically connected by currents (Nanninga et al. 2015). Then environmental differences, like difference in habitat, might be preventing the colonization of new recruits (Giles et al. 2015), which also could be occurring in the ARMS community. In this work, the taxa Echinodermata and Palaemonidae decreased in abundance from inshore to offshore and the Galatheidae showed an inverse relationship. Having information on biodiversity will help design better MPAs (Ward et al. 1999; Cowling et al. 2003). For MPAs with the goal of preserving biodiversity, protecting reefs with high biodiversity and the reefs exporting larvae to high biodiversity reefs would be a priority. Here Al Fahal South had the greatest biodiversity.

The communities of the ARMS on the reefs near KAUST are dominated by a few species, with 70% and 45% of species on the barcoded and metabarcoded fractions, respectively, being rare species (with less than 3 sequences). The evenness of all the samples in this study was low, with a steep slope on the rank abundance plot (Fig. 4). This pattern had been observed on other coral reefs and seems to be a constant in ARMS communities of reefs (Plaisance et al. 2011a; Leray & Knowlton 2015). It had been suggested that rare species may be arriving to a community through migration (Hubbell 2001; Magurran & Henderson 2003). Magurran and Henderson (2003) predict that a considerable change in the
conditions will make a rare species take the place of an affected common species. Further studies comparing communities on the Red Sea, before and after a catastrophic event for the cryptic community and observing for shifts on the common species could elucidate this for the ARMS community.

The results suggest that Fsar (an inshore reef) has a different community in the barcoded fraction than the other reefs. The inshore reef also has a different community composition in benthic and fish surveys (Furby et al. 2013; Khalil 2015). Khalil (2015) argued that inshore reefs difference in community composition as compared to midshelf and offshore reefs could be due to the 2010 bleaching event, which affected the inshore reefs to greater extent (Furby et al. 2013). We retrieved the ARMS five years after the latest bleaching event at that time, suggesting that ARMS communities recover at a low pace or there are other factors affecting the distribution of the species on inshore reefs.
Chapter 5: Conclusions

The biodiversity of the Red Sea is considerably understudied and even with metabarcoding techniques and standardized sampling methods, more exhaustive sampling is needed to discover the full extent of biodiversity in this system. Here 10% and <1% of the OTUs of the barcoded and metabarcoded fractions respectively got a BLAST hit with a previously identified species on the NCBI database. In addition none of the rarefaction curves plateau. The inshore reefs have a different community on the ARMS than the midshelf and offshore reefs. The difference between midshelf and offshore reefs on the ARMS community is not clearly defined. The species diversity and abundance of the ARMS community seem to be correlated to changes on microhabitat, considering the encrusting organisms as microhabitats for the cryptic fauna. Algae and bryozoan cover were the microhabitats that showed a significant relationship with differences on community structure. Sediment concentration on the ARMS did not explain the differences in communities as it was expected. By revealing biodiversity patterns and understanding what drives the changes in communities this work for marine reserve planning on the region.
REFERENCES


22. 

23. 

24. 

25. 

26. 

27. 

28. 

29. 

30. 

31. 

32. 

34.


35.


36.


37.


38.


39.


40.


41.


42.


43.


44.


